

**National Health and Nutrition Examination Survey:  
Analytic Guidelines, 2011-2012**

September 30, 2013

Centers for Disease Control and Prevention  
National Center for Health Statistics  
Division of Health and Nutrition Examination Surveys

## Table of Contents

<b>1. Introduction.....</b>	<b>3</b>
<b>2. Data Considerations.....</b>	<b>4</b>
2.1. Sample design changes for the 2011-2014 sample .....	4
2.2. Key variables of interest and impact of disclosure assessment .....	7
2.3. Effect of nonresponse .....	9
<b>3. Analytic Considerations .....</b>	<b>11</b>
3.1 Age adjustment .....	11
3.2 Computing population counts .....	12
<b>4. Conclusion .....</b>	<b>13</b>
<b>5. Reference .....</b>	<b>13</b>

## Text Tables

- A. Unweighted sample size and percentage by race and Hispanic origin for examined participants, NHANES 2009-2010 and 2011-2012
- B. Differences in the age-related variables on public data files, NHANES 2009-2010 and 2011-2012
- C. Overall unweighted survey response rates for all ages, NHANES 1999-2012

## 1. Introduction

This report presents information regarding the analytic and reporting guidelines for the 2011-2012 NHANES publicly released data. This document is an addendum to the *National Health and Nutrition Examination Survey (NHANES): Analytic Guidelines, 1999-2010*, hereafter referred to as the 1999-2010 Analytic Guidelines (available at: [http://www.cdc.gov/nchs/nhanes/nhanes2003-2004/analytical\\_guidelines.htm](http://www.cdc.gov/nchs/nhanes/nhanes2003-2004/analytical_guidelines.htm)). All the guidelines provided in that document are relevant to data from the 2011-2012 survey cycle, so users are urged to review that document before reading the sections below.

Note that the statistical guidelines in this document, and the 1999-2010 Analytic Guidelines, are not standards. Depending on subject matter and statistical efficiency, specific analyses may depart from these guidelines. In conducting analyses, the analyst needs to use his or her subject matter knowledge (including knowledge of methodological issues), as well as information about the survey design. The more an analyst deviates from the original analytic categories defined in the sample design, the more important it is to evaluate the results carefully and to interpret the findings cautiously.

The recommended approach for analysis of NHANES data is design-based analysis. Design-based analytic procedures explicitly take into account features of the survey design such as differential selection probabilities and geographic clustering. The 2011-2012 survey cycle is part of a four-year sample including data collected in 2013 and 2014. Therefore, this document will provide specific information on the 2011-2012 survey cycle, but also on the 2011-2014 sample, when relevant. It is important to note that data from a two year cycle, such as 2011-2012, may not meet all analytic objectives of the full four year sample and thus should be considered only a preliminary snapshot of the four year NHANES sample.

We strongly encourage that all data users, prior to any analysis of NHANES data, read all relevant documentation on the survey overall and for the specific data files to be used in their analysis. Specific data file documentation can be found via the link next to the respective data file on the NHANES website. An additional resource for all analysts is the series of NHANES Tutorials (<http://www.cdc.gov/nchs/tutorials/index.htm>) — a Web-based product designed to assist users in understanding and analyzing NHANES data.

## **2. Data Considerations**

The sample design used since 1999 allows the production of aggregate-level national estimates from NHANES each year from a multi-year sample design. However, while annual samples are nationally representative, estimates for single year data are relatively unstable (have large variance estimates) since NHANES can only go to a small number of primary sampling units (PSUs) each year. In addition, releasing only one year of data increases the possibility of disclosure of a sample person's identity. This along with the analytic limitations of the annual sample resulted in the decision to publicly release data in two-year cycles and to keep the survey content within those years fixed to the extent possible.

### **2.1. Sample design changes for the 2011-2014 sample**

Over the years the NHANES design has changed to sample larger numbers of certain subgroups of particular public health interest. Oversampling is done to increase the reliability and precision of estimates of health status indicators for these population subgroups. Weighting schemes allow estimates from these subgroups to be combined to obtain national estimates that reflect the relative proportions of these groups in the population as a whole.

For NHANES 2007-2010 the sample design was changed to oversample all Hispanic persons rather than just Mexican-American Hispanic persons. This was done in a way which resulted in sample sizes sufficient to produce reliable estimates for Mexican Americans in addition to Hispanics overall. A key change between NHANES 2007-2010 and NHANES 2011-2014 was that Asians were also oversampled in addition to the ongoing oversample of Hispanics, non-Hispanic blacks, older adults, and low income white and other persons.

Specifically, the over-sampled subgroups in 2011-2012 survey were as follows:

- Hispanic persons
- Non-Hispanic black persons
- Non-Hispanic Asian persons
- Non-Hispanic white and Other persons at or below 130 percent of the federal poverty level
- Non-Hispanic white and Other persons aged 80 years and over

The Asian category includes all persons having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent, including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

In order to meet target sample design specifications for Asians, Hispanics, and black persons, it was necessary to make these three groups mutually exclusive. This was accomplished by using the above specified categories. So, any Asian person who also reported being Hispanic or non-Hispanic black was considered to be in the respective latter categories, since there were separate sampling rates for those groups. A seemingly more efficient approach could have been to sample all Asians into that category, regardless of their Hispanic origin or whether they also self-identified as black. However, since only 1.7 percent of all Asians also self-identify as black, and only 0.3 percent of Hispanics self-identify as Asian<sup>1</sup>, little efficiency was lost by restricting the Asian oversample to non-Hispanic, non-black Asians.

To reflect the change in the sample design for this survey cycle, an additional race/Hispanic origin variable, RIDRETH3, is included on the 2011-2012 public-use Demographic data file. This variable is consistent with the race/Hispanic origin variable, RIDRETH1, which is available on previous survey data releases in that the Mexican American and Other Hispanic categories may include persons of multiple races and the non-Hispanic white, non-Hispanic black and non-Hispanic Asian categories include only those reporting a single race. All non-Hispanic persons reporting multiple races are in the “Other race” category. Table A shows the 2011-2012 sample distribution by this new variable compared to the previous cycle. RIDRETH1 is still included on the 2011-2012 Demographic data file so that comparison of NHANES 2011-2012 with previous survey cycles (without the Asian oversample) is still possible.

While the weighted distribution of the sample across the four major race/Hispanic origin categories is controlled to the U.S. distribution of these groups, it is not controlled to the distribution of the individual subgroups within a race/Hispanic origin group, such as Asians.

**Table A. Unweighted sample sizes and percentages by race and Hispanic origin for examined participants, NHANES 2009-2010 and 2011-2012**

Survey years	Hispanic		Non-Hispanic				Total
	Mexican American	Other Hispanic	White, single race	Black, single race	Asian, single race	Other race <sup>2</sup> , including multiracial persons	
2009-2010 n (%)	2,305 (22.5)	1,103 (10.8)	4,317 (42.1)	1,903 (18.6)	n/a <sup>1</sup>	625 (6.1)	10,253 (100)
2011-2012 n (%)	1,316 (14.1)	1,011 (10.8)	2,841 (30.4)	2,582 (27.7)	1,215 (13.0)	373 (4.0)	9,338 (100)

<sup>1</sup>Non-Hispanic Asians were included in the “Other” category in 2009-2010.

<sup>2</sup>Other race includes non-Hispanic persons who reported a race other than white, black, or Asian or reported more than one race.

Since the total sample size in any year is fixed due to operational constraints the increase in the Asian sample size resulted in a decrease in the percent examined for non-Hispanic whites, Mexican Americans, and Other race groups. For the NHANES 2011-2012 sample, Asians and Mexican Americans each comprise about 13-14 percent of the unweighted examined sample. This limited sample size means that detailed two and three-way analytic comparisons of demographic subgroups will likely not meet all analytic criteria as outlined in the 1999-2010 Analytic Guidelines. Therefore, data users are cautioned to review sample sizes prior to attempting analyses previously performed by race/Hispanic origin groups. With small sample sizes, analysts should consider combining subgroups or to postpone analyses until the release of data from the NHANES 2013-2014 survey cycle, which will also contain an oversample of Asians.

The oversampling of subgroups mentioned above, and the operational variability across PSUs, can cause the NHANES weights to be quite variable. Further, when subdomains are combined for analysis, a wide range of weights may occur, which will lead to increased variance in the analytic results. This will particularly be the case when combining 2011-2012 data for Asians with other groups, since the Asian sample has much lower weights than other groups. Analysts should be aware of the range of weights within the subgroup being analyzed and the resulting potential increase in variance. Analysts should also be aware of the potential influence that cases with large weights can

have on their analyses, especially when extreme weights are associated with extreme data points. Note that the largest sample weights in the 2011-2012 survey cycle data are larger than the largest weights in previous years.

## **2.2. Key variables of interest and impact of disclosure assessment**

NHANES data collection adheres to the requirements of Federal Law. The Public Health Service Act (42 USC 242k) authorizes data collection and Section 308(d) of that law (42 USC 242m), the Privacy Act of 1974 (5 USC 552A), and the Confidential Information Protection and Statistical Efficiency Act (PL 107-347) prohibit NCHS from releasing information that may identify any respondent or group of respondents. As a result, data edits must be made to some variables to address data disclosure concerns.

With the addition of the Asian oversample, and the public release of a more detailed race and Hispanic origin variable, additional edits were necessary to variables previously released with earlier survey cycles in order to maintain the confidentiality pledge made to NHANES respondents. These increased edits can be seen in the variables indicating the respondent's age and country of birth, among others.

Similar to previous data release cycles, the 2011-2012 demographic file includes a variable for age in years at screening (RIDAGEYR) for all participants. A new variable has been added to denote the age in years at examination (RIDEXAGY) for participants aged 2 to 19 years. Age in months at screening (RIDAGEMN) and at examination (RIDEXAGEM) are reported for participants aged 0 to 24 months. Due to increasing concerns about potential disclosure risks, information on age in months at screening and at examination for participants in other age groups are no longer included in the public release file but are available through the [NCHS Research Data Center \(RDC\)](#). The differences in the age-related variables available on the 2009-2010 and 2011-2012 publicly released data files are illustrated in Table B.

**Table B. Differences in the age-related variables on public data files, NHANES 2009-2010 and 2011-2012**

<u>Variable name</u>	<u>Description</u>	<u>2009-2010 data file</u>	<u>2011-2012 data file</u>
RIDAGEYR	Age in years at screening (for persons aged 0-80 years)	Yes	Yes
RIDAGEMN	Age in months at screening (for persons aged 0-80 years)	Yes	Yes - for children 24 months or younger
RIDAGEEX	Age in months at MEC examination (for persons aged 0-80 years)	Yes	No - replaced with RIDEXAGY and RIDEXAGM for those 19 years or younger
RIDEXAGY	Age in years at MEC examination (for persons aged 2-19 years at screening)	No	Yes
RIDEXAGM	Age in months at MEC examination (for persons aged 0-24 months at screening)	No	Yes

Age at screening was used to determine eligibility for an examination component and should be used for most analyses. However, in prior years, when analyzing anthropometric data on children and youth from birth through 19 years, age in months at MEC examination (RIDAGEEX) was the recommended age variable for analyses. Given the changes in the age-related variable on the 2011-2012 public data files, a new variable was created, BMDBMIC, which provides analysts pre-computed BMI categories for children and adolescents aged 2 to 19 years at examination. For further details refer to the Body Measures Data File and Documentation.

In 2009-2010, the variable DMDBORN2 was available on the publicly released data and included categories of “Mexico,” “Other Spanish Speaking Country,” and “Other Non-Spanish Speaking Country.” In 2011-2012, the variable DMDBORN4 is available and has only two categories: “Born in 50 US States or Washington, DC,” and “Others.”

Refer to the NHANES 2011-2012 Demographic Data Documentation for further details on these and other demographic variables which may have been released differently in previous years. Analysts are again reminded to carefully review the data file documentation before starting any analysis.



### 2.3. Effect of nonresponse

Not all persons selected in the NHANES sample were interviewed and not all interviewed persons were examined. Unit or participant nonresponse, the failure to obtain any information on an individual selected to participate in an NHANES survey, can occur both at the interview and at the examination phase of the survey. Nonresponse bias resulting from this missing data can be an important source of survey error, and can be substantial when two conditions hold: 1) the response rate is relatively low and 2) the difference between the characteristics of respondents and nonrespondents is relatively large.

In 2011-2012, 13,431 persons were selected for NHANES from 30 different study locations. Of those selected, 9,756 completed the interview and 9,338 were examined. This resulted in an overall response rate lower than previously experienced in recent years of NHANES, as shown in Table C. Detailed response rates by age, sex, and race/Hispanic origin are available at [http://www.cdc.gov/nchs/nhanes/response\\_rates\\_cps.htm](http://www.cdc.gov/nchs/nhanes/response_rates_cps.htm)). Response rates were lowest for Asians and those 60 years old and older. Given this, a nonresponse bias analysis was performed.

**Table C. Overall unweighted survey response rates for all ages, NHANES 1999-2012**

Survey years	Screened sample <sup>1</sup>	Interviewed sample		Examined sample	
		Sample size	Response rate (percent)	Sample size	Response rate (percent)
1999–2000	12,160	9,965	81.9	9,282	76.3
2001–2002	13,156	11,039	83.9	10,477	79.6
2003–2004	12,761	10,122	79.3	9,643	75.6
2005–2006	12,862	10,348	80.5	9,950	77.4
2007–2008	12,943	10,149	78.4	9,762	75.4
2009–2010	13,272	10,537	79.4	10,253	77.2
2011–2012	13,431	9,756	72.6	9,338	69.5

<sup>1</sup> Screener response rates across survey cycles have varied from 98-99% and the loss of eligible respondents at this stage is negligible.

The nonresponse bias analysis was performed for the overall population in addition to Asians, elderly (age 60 years and older), and young children (age 1 to 5). The analysis was conducted in two stages. An initial analysis involved the comparison of demographic and socio-economic characteristics of respondents to those of nonrespondents. Further analyses were then conducted focusing on nonresponse bias in final outcome statistics. All analyses used design based methods for estimating variances to the extent possible.

The initial analyses showed some indication of potential for bias in the respondent sample, prior to conducting nonresponse adjustments, to the extent that the characteristics analyzed are related to health. However, several of the characteristics found to be significantly related to response status were either used or highly correlated with those used in the weighting adjustments indicating that the bias may have been reduced through the weighting adjustments.

To determine if any of the potential bias identified in the analyses described above remained after the weighting adjustments, estimates of the characteristics of selected persons from the full sample (including the nonrespondents) were compared to estimates for the respondents only before and after weighting adjustments for interviewed and examined persons. The results showed that very few estimates indicated large relative differences across the stages of weighting for interviewed persons, and none indicated large relative differences for examined persons.

Estimates from the NHANES 2011-2012 survey cycle were compared with comparable estimates from the National Health Interview Survey (NHIS). In general, the differences between NHIS and NHANES for most of the estimates were small. Estimates from the NHANES 2011-2012 survey cycle were also compared with the estimates from previous survey cycles. Some significant differences were found for laboratory data, as the result of changes in laboratories and methodologies, and some body measurements, which may be due to actual changes in those data over time.

The nonresponse bias analyses performed to date demonstrate potential nonresponse bias before weighting adjustments, but that the weighting adjustments reduced this bias. The methods and results of these and other nonresponse bias analyses will be presented in

detail in a forthcoming report.

### 3. Analytic Considerations

The most important considerations in analyzing NHANES data involve taking into account the survey design. Survey sample weights should be used and the complex survey design must be accounted for in the estimation of variance. These weights account for oversampling and survey nonresponse, and their proper usage ensures that calculated estimates are truly representative of the U.S. civilian non-institutionalized population.

Again, the user is directed to the 1999-2010 Analytic Guidelines. That document provides guidance on how to determine the appropriate survey sample weights for analyses, combining two-year weights to analyze multi-year samples, variance estimation and appropriate procedures for subsetting the NHANES data. Information on age adjustment and computing population counts are also provided, but will be further discussed here as it relates specifically to the 2011-2012 data.

#### 3.1 Age adjustment

Age-adjustment is important to consider for trend analyses between NHANES cycles and for comparisons between subgroups with different age distributions within NHANES as recommended in the 1999-2010 Analytic Guidelines. For the NHANES 2011-2012 survey cycle, the **recommended standard population continues to be the year 2000 population projections from the U.S. Census Bureau<sup>2</sup>**.

The following standard proportions are based on the 2000 standard population and should be used in NHANES 2011-2012 analyses when using 20-year age groups for ages 20 years and over.

<i>Age Group</i>	<i>Proportion</i>
20-39	0.3966
40-59	0.3718
60 and over	0.2316

Prior to NHANES III, the NHES and NHANES had upper age limits so trends need to be conducted on ages 20-74 years. Consequently, to compare age-adjusted estimates for

NHANES 1999-2012 with these surveys, the following standard proportions should be used:

<i>Age Group</i>	<i>Proportion</i>
20-39	0.4332
40-59	0.4062
60-74	0.1606

Any comparison of age-adjusted rates requires that the same standard population and the same age groups be used. For example, it is not appropriate to compare an age-adjusted rate from NHANES III based on the 1990 standard to an age-adjusted rate from NHANES 2011-2012 based on the 2000 standard.

### **3.2 Computing population counts**

To understand the public health impact of a condition, it is often helpful to calculate population counts in addition to the prevalence of a health condition. By quantifying the number of people with a particular condition or risk factor, counts speak directly to the burden or magnitude.

Since NHANES is a nationally representative survey of the civilian noninstitutionalized U.S. population, population estimates are based on reliable estimates for this aspect of the U.S. population. For computing estimates for NHANES 1999-2010, totals from the Current Population Survey (CPS) for each survey cycle were made available by race and Hispanic origin, sex and age. For NHANES 2011-2012, the totals provided are from the 2011 American Community Survey (ACS). This change was made in part as a result of the addition of the Asian oversample. With this addition, it became necessary to ensure that the totals provided reliable estimates of the total of Asians within age and sex categories. While both the CPS and ACS are surveys, the sample size for the ACS is about 13 times that of the CPS. This larger sample size resulted in more reliable estimates for the Asian population. Population totals are available at [http://www.cdc.gov/nchs/nhanes/response\\_rates\\_CPS.htm](http://www.cdc.gov/nchs/nhanes/response_rates_CPS.htm)

The calculation of population counts is described in the 1999-2010 Analytic Guidelines. It also is possible to combine NHANES survey cycles (See section 2.1 for information on changes in the NHANES 2011-2012 sample design affecting estimates made from combined survey cycles including the 2011-2012 data). Even though the source of

population estimates for the 2011-2012 survey cycle changed from previous cycles, population estimates may still be combined across these cycles in the usual manner. For example, to combine 2011-2012 and 2009-2010, combine them as follows in order to get a population total for 2009-2012:  $\frac{1}{2}$  (NHANES 2011-2012 population totals) +  $\frac{1}{2}$  (NHANES 2009-2010 population totals).

#### **4. Conclusion**

In summary, these analytic guidelines represent the latest statistical procedures and analytic guidance for the continuous NHANES survey for the years 2011-2012. Any significant changes in the NHANES survey design or the introduction of new statistical techniques for the analysis of complex sample surveys will result in a revised version of these guidelines.

#### **5. Reference**

1. U.S. Census Bureau, 2011 American Community Survey
2. Klein RJ, Schoenborn, CA. Age Adjustment Using the 2000 Projected U.S. Population. Healthy People Statistical Notes, no 20. Hyattsville, Maryland: National Center for Health Statistics. January 2001.